INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-26

An Attempt at Reading Keynes's Treatise on Probability

C. P. ROBERT[1]

[1] Université Paris-Dauphine, CEREMADE and CREST-INSEE, Paris.

An attempt at reading Keynes's Treatise on Probability

Christian P. Robert
Université Paris-Dauphine, CEREMADE, and CREST, INSEE

Abstract. The book *A Treatise on Probability* was published by John Maynard Keynes in 1921. It contains a critical assessment of the foundations of probability and of the current statistical methodology. As a modern reader, we review here the aspects that are most related with statistics, avoiding a neophyte's perspective on the philosophical issues. In particular, the book is quite critical of the Bayesian approach and we examine the arguments provided by Keynes, as well as the alternative he proposes. This review does not subsume the scholarly study of Aldrich (2008a) relating Keynes with the statistics community of the time.

Keywords: Daniel Bernoulli, Ronald Fisher, Pierre Simon de Laplace, Whilhem Lexis, Harold Jeffreys, Karl Pearson, probability theory, frequency, Law of Large Numbers, foundations, Bayesian statistics, history of statistics.

1 Introduction

> *" A definition of* probability *is not possible, unless it contents us to define degrees of the probability-relation by reference to degrees of rational belief."*
> **A Treatise on Probability**, *page 8.*

Following an earlier and fruitful reading seminar on Jeffreys' *Theory of Probability*, that led to a critical assessment in Robert et al. (2009), and based on a side remark in Taleb (2008) as to the importance of the book[1], I endeavoured to assess Keynes (1921) in preparation for my 2010 graduate reading seminar.

A Treatise on Probability is John Maynard Keynes' polishing of his 1907 and 1912 Fellowship dissertation into a book after an interruption due to the war. The proclaimed ambitious goals of establishing a logical basis for probability and of drawing a new approach for statistical induction and the extremely strong views contained of the book, as well as the highly critical reassessments of most authors, are reflecting on the youth of the author, who was born in 1883 and was therefore 24 when he drafted *A Treatise on Probability*. The extensive coverage of the literature of the time and the thorough discussion of the various theories in competition shows the extent of the scholarly approach of Keynes to the fields of probability and statistics, but the methodological innovation contained in the book is somehow limited, in line with both the title and Keynes' acknowledgement that he is *"unlikely to get much further"*. Keynes'view

[1] *"Treatise on Probability remains in many people's opinion the most important single work on the subject"* (Taleb, 2008).

on statistics are missing some of the latest developments of the time, maybe due to his anti-probabilist stance. The statistical novelty in *A Treatise on Probability* mostly concentrates on Part V, Chapter 33, remaining mostly at a rethorical level and wishing for prospective followers that never materialised. The conclusion of the exercise is therefore that the book has a restricted methodological (rather than historical and philosophical) interest and this—along with the fact that Ronald Fisher was about to publish two of his most influential papers—explains why Keynes' incursion in probability and statistics did not have a lasting impact, even those most sympathetic to the book (Jeffreys, 1931; Lindley, 1968) seeing no practical nor methodological aspect to draw from. Stigler (2002, pp. 161-162) also discusses the limited worth of *A Treatise on Probability* as a mathematical and statistical treatise, with almost sole focus *"the binomial world"* and unable to *"carry the weight of a serious social scientific investigation"*.

In this note, I cover what I consider to be the most interesting parts of the book, on connection with current Statistics, avoiding the seemingly outdated philosophical debates about the nature of probability and of induction that constitute most of the book and do not overlap with my own interests.[2] I want to stress the fact that (Aldrich, 2008a) published an extensive and scholarly coverage of the impact (or lack thereof) of Keynes and of *A Treatise on Probability* on the philosophical and statistical communities at the time. I only became aware of this impressive survey when starting to compose this note and it considerably helped me to anchor the impressions I had about the book within the intellectual scene of the time. As in Robert et al. (2009), there is no attempt at an history of statistics in this review, which is to be taken as a reflection of a 2010 Bayesian reader upon a piece of work written one century ago. For historical aspects on inverse probability, I refer the reader to the comprehensive coverage in Dale (1999) who, ironically, stops its range at Karl Pearson, i.e. just before Keynes attempted to enter the statistical scene.

2 A restricted perspective

> *"Statistical techniques tell us how to 'count the cases'."* **A Treatise on Probability**, *page 392.*

The most interesting feature of the book, from a statistician's viewpoint, is the weakness of the statistical principles advanced therein! Obviously, this point cannot be construed as a criticism of Keynes if only because the title of the book, as the memoir's intention is to set probability within logic, rather than with mathematics. Nonetheless, given the contents of Jeffreys' (1939) and de Finetti's (1937, reprinted as 1974) homonymous *Theory of Probability*, it could have been a possibility that the book would provide a more thorough treatment of the theory of Statistics of the time, if not proposing advances in this domain. This was actually my purpose when I started reading the book.

[2]Readers interested in the philosophy of probability could however relate the first parts of the book with Burdzy's (2009) recent manifesto since both authors seem to be debating on the same plane.

For instance, when compared with the much more modern treatise by Jeffreys (1931)[3], this book does not contain analyses of realistic datasets, except when criticising von Bortkiewicz's theory through the Prussian cavalry horsekick data. It seems that Keynes spend a large portion of his manuscript in decrying a large part of the current statistical practice (first and foremost, Bayesian[4] statistics!) as well as a majority of the past and current statisticians, and in reproducing the arguments of other researchers, like Boole, Lexis, or von Kries. (Again, this is in line with the argument that the book is a scholarly and critical memoir rather than a innovative manifesto, even though the author aimed at a broader impact on the statistics community at the time.) Furthermore, most of Part V, entitled *Statistical Inference* and the centre of focus in the present analysis, deals with observation frequencies and their stabilisation. This is somehow surprising when considering the main research field of Keynes, namely Economics, where examples abound. Instead, a very small number of (academic) examples like the proportion of boys in births is recurrently discussed throughout the book.

To be complete about the statistical contents of *A Treatise on Probability*, Part II on *Fundamental Theorems* also contains a chapter on the properties of various estimators of the mean in connection with the distribution of the observations, although Keynes dismisses its importance by stating *"It is without philosophical interest and should probably be omitted by most readers"* (page 186). It actually reproduces Keynes' only statistical paper, published in the *Journal of the Royal Statistical Society* in 1911 on the theory of averages (to be discussed in Section 2.3).

2.1 The role of models

> *"It seldom happens that we can apply Bernoulli's theorem to a long series of natural events."* **A Treatise on Probability**, *page 343.*

Throughout the book, Keynes seems to hold (probability as a mathematical theory and) models in very low regard, considering that unknown probability do not exist and that reproducibility is almost always questionable (*'Some statistical frequencies are, with narrower or wider limits, stable. But stable frequencies are not very common, and cannot be assumed lightly"*, page 336), apart from urn models. He thus falls into what we could call the "ultra-conditioning fallacy", namely that, the more covariates

[3]Jeffreys (1922) reviewed the book for *Nature*. His review was quite benevolent, despite most of Keynes' perspectives on statistics being foreign to his own. This may explain why Part V was mostly bypassed by the review. His comments in *Scientific Inference* (1931, pp. 222-224) about Keynes' refusal to admit that probabilities were numbers, hence comparable, and in *Theory of Probability* (1939, p. 25) about Keynes' unwillingness to generalise axioms, more truthfully reflects Jeffreys' global opinion about the book.

[4]It is worth pointing out that the denomination appeared much later (Fienberg, 2006), Keynes resorting to the then current denomination of Inverse Probability or, in a more derogatory way, *"the Laplacian theory of 'unknown probabilities"* (page 372). According to (Aldrich, 2008a), the early papers of Keynes on statistics were definitely Bayesian with most of his analysis being based on an uniform prior. He later started to worry about the influence of the prior, leading to *A Treatise on Probability* and its harsh criticism of inverse probability, even though some Bayesian arguments remain in use within the book.

you condition upon, the more different the individuals behave, a point of view that goes against supporting statistical practice. For instance, Keynes states that *"where general statistics are available, the numerical probability which might be derived from them is inapplicable because of the presence of additional knowledge with regards to the particular case"* (page 29). (He then goes on deriding Gibbon for his use of mortality tables when he should have called for a doctor!) This shows the gap between the perspectives of Keynes and those of Jeffreys (1931, 1939) and de Finetti (1937, reprinted as 1974), the later focussing on the exchangeability of events to derive the existence of a common if unknown probability distribution.

The use of particular sampling distributions (called *laws of errors*) in his reproduction of his 1911 paper on averages is not discussed in a modelling perspective but simply to back up the standard types of averages as maximum likelihood estimators.[5] *"The general evidence which justifies our assumption of the particular law of errors which we do assume"* (page 195) is never discussed further. Model choice was not an issue in 1920!

2.2 Keynes as a critical frequentist

> *"The frequency theory, therefore, entirely fails to explain or justify the most important source of the most usual arguments in the field of probable inference."* **A Treatise on Probability**, *page 108.*

Given the above mentioned reluctance to accept models and reproducibility, it is no surprise that only extensive frequency stability is acceptable for Keynes: *"The 'Law of Large Numbers' is not at all a good name for the principle that underlies Statistical Induction. The 'Stability of Statistical Frequencies' would be a much better name for it."* (page 336).

> *"Some statistical frequencies are, with narrower or wider limits, stable. But stable frequencies are not very common, and cannot be assumed lightly."* **A Treatise on Probability**, *page 336.*

The criticism of the Central Limit Theorem (CLT), called Bernoulli's Theorem in the book,[6] that is found in Chapter XXIX is rather curious, in that it confuses model probabilities p with probability estimates p' (the identical notation being an indicator of

[5] Aldrich (2008b) argues quite convincingly that these are rather maximum a posteriori (MAP) estimates corresponding to flat priors, making the term *most probable* more coherent. When stating the problem on page 194, Keynes indeed invokes the *"theorem of inverse probability"*, namely Bayes' theorem on the parameter *finite* set. Again, this shows that part of Keynes' reasoning is still grounded within the principles of inverse probability.

[6] Bernoulli's Theorem is historically the weak Law of Large Numbers but Keynes presents this result in conjunction with (a) a description of the binomial $\mathcal{B}(n,p)$ distribution and (b) the normal (CLT) approximation to the binomial cdf, a result he calls Stirling's theorem. While Edgeworth had a clear influence on Keynes, his expansions providing a better approximation are not mentioned.

this confusion).[7] For instance, on page 343, Keynes criticises the use of the Central Limit Theorem (CLT) for the Bernoulli distribution $\mathcal{B}(1/2)$ and a coin tossing experiment as, when *"heads fall at every one of the first 999 tosses, it becomes reasonable to estimate the probability of heads at much more than* $1/2$*"*. This argument is therefore confusing the probability model $\mathcal{B}(p)$ with the estimation problem. Keynes' inability to recognise the distinction may stem from his reluctance to use unknown probabilities such as p. Similarly, on pages 349-350, when considering the proportion of male births, Keynes states that the probability of having n males births in a row is not p^n, if p is the probability of a single male birth, but

$$\frac{r}{s} \cdot \frac{r+1}{s+1} \cdot \frac{r+2}{s+2} \cdots \frac{r+n-1}{s+n-1}$$

if s is the number of births observed so far, and r the number of male births. The later is a sequential construction based on individual estimates for each new observation, not a true (predictive) probability. In fact, under a flat prior on p, the predictive (marginal) probability of seeing n male births in a row is

$$\int_0^1 p^n \frac{(s+1)!}{r!(s-r)!} p^r (1-p)^{s-r} \, \mathrm{d}p = \frac{(s+1)!}{r!(s-r)!} \frac{(r+n)!(s-r)!}{(n+s+1)!} = \frac{(r+1)\cdots(r+n)}{(s+2)\ldots(s+n+1)} \,.$$

Keynes' solution actually corresponds to using Haldane's prior, $\pi(p) = 1/(p(1-p)$ (see Jeffreys, 1939).

To conclude that Bernoulli's Theorem (a simple version of the CLT in the binomial case) does not hold exactly in this setting is thus inappropriate. When considering that *"knowledge of the result of one trial is capable of influencing the probability at the next"*, the "true" probability is confused with the estimated one. The same criticism applies to the remark that *"a knowledge of some members of a population may give us a clue to the general character of the population in question."* (page 346), remark that bears witness to Keynes' skepticism about the relevance of probabilistic models. From a Bayesian perspective, it could be said that Keynes mixes sampling distributions with marginal distributions, as in the latent variable example of page 346 dealing with observations from $\mathcal{B}(p)$ when $p \in \{p_1, \ldots, p_k\}$: the observations become dependent when integrating out p. The sentence *"if we knew the real value of the quantity, the different measurements of it would be independent"* (page 195) may be understood under this light, even though it is a risky extrapolation given the book stance on Bayesian statistics and the lack of evidence Keynes mastered this type of integration subtleties.

2.3 Keynes's views on current statistical inference

> *"The statistician is less concerned to discover the precise conditions in which a description can be legitimately extended by induction."* **A Treatise on Probability**, *page 327.*

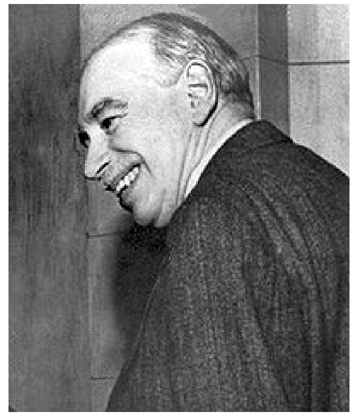
[7] Jeffreys (1931, p. 224) stresses that *"Keynes' postulate might fit the assigned probabilities instead of the true probabilities"*.

Chapter XVII reproduces Keynes' 1911 paper on the distributions leading to specific averages as MLEs, then obtaining classes of densities for which the arithmetic, the geometric and the harmonic averages as well as the median are MLES. The decision-theoretic justifications of the arithmetic mean by Laplace and Gauß are derided as depending on *"doubtful and arbitrary assumptions"* (page 206), while the lack of reparameterisation invariance of the arithmetic average as MLE is clearly stated (on page 208).

The focus of statistical inference as described in Part V is reduced to probability assessment: *"In the first type of argument we seek to infer an unknown statistical frequency from an* à priori *probability. In the second type we are engaged on the inverse operation, and seek to base the calculation of a probability on an observed statistical frequency. In the second type we seek to pass from an observed statistical frequency, not merely to the probability of an individual occurrence, but to the probable value of other unknown statistical frequencies"* (page 331). This is actually rather surprising given the overall negative tone of *A Treatise on Probability* about probability theory.

The first item is actually a probabilistic issue and is treated as such in Chapter XXIX, which covers the normal and Poisson limiting theorems, as well as Čebyšev's inequality. The criticisms of "Bernoulli's Theorem" found in this chapter are limited to the fact that finding independent and identically distributed (i.i.d.) replications is a condition that is *"seldom fulfilled"* (page 342). The 1901 proof by Liapounov of the CLT for general independent random variables is not mentioned in Keynes' book and was presumably unknown to the author. Instead, he refers to Poisson for a series of independent random variables with different distributions, warning that *"it is important not to exaggerate the degree to which Poisson's method has extended the application of Bernoulli's results"* (page 346). Although Čebyšev's inequality has very little impact on statistical practice, except when constructing conservative confidence intervals, Keynes is clearly impressed by the result (of which he provides a very convoluted proof on pages 353-355) and he concludes that *"Laplacian mathematics is really obsolete and should be replaced by the very beautiful work which we owe to these Russians"* (page 355). Chapter XXIX terminates with an interesting section on simulation experiments aiming at an empirical verification of the CLT, although Keynes' conclusion on a very long dice experiment is that, given that the frequencies do not match up *"what theory would predict"* (page 363), the dice used in this experiment was quite irregular (or maybe worn out by the 20,000 tosses!)

> *"I do not believe that there is any direct and simple method by which we can make the transition from an observed numerical frequency to a numerical measure of probability."* **A Treatise on Probability**, *page 367.*



As discussed in the next section, Keynes does not consider Laplace's (i.e. the Bayesian) approach to be logically valid and he similarly criticises normal approximations à la Bernoulli, seeing both as *"mathematical charlatanery"* (page 367)! Even the (maximum likelihood) solution of estimating p with the frequency x/n when $x \sim \mathcal{B}(n,p)$ does not satisfy him (as being *"incapable of a proof"*, page 371). Note that the maximum likelihood estimator is called the *most probable value* throughout the book, in concordance with the current denomination at the time (Hald, 1999), without Keynes objecting to its Bayesian flavour.[8] Obviously, given that he wrote the main part of the book before the war, he could not have used Fisher's denomination of maximum likelihood estimation since its introduction dates from 1922.[9]

The recovery in Chapter XVII of standard averages as MLEs is more of a technical interest than of true methodological relevance, because the classification of distributions (*"laws"*) that give the arithmetic, geometric, harmonic mean or the median as MLEs is obviously parameterisation-dependent, a fact later noted by Keynes but omitted at this stage despite his criticism of Laplace's principle on the same ground. The derivation of the densities $f(x,\theta)$ of the distributions is based on the condition that the likelihood equation

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(y_i,\theta) = 0$$

is satisfied for one of the four empirical averages, using differential calculus despite the fact that Keyne earlier derived on page 194 Bayes' theorem by assuming the parameter space to be discrete.[10] Under regularity assumptions, in the case of the arithmetic mean, this leads to the family of distributions

$$f(x,\theta) = \exp\left\{\phi'(\theta)(x-\theta) - \phi(\theta) + \psi(x)\right\} ,$$

where ϕ and ψ are arbitrary functions such that ϕ is twice differentiable and $f(x,\theta)$ is a density in x, meaning that

$$\phi(\theta) = \log \int \exp\left\{\phi'(\theta)(x-\theta) + \psi(x)\right\} \mathrm{d}x ,$$

a constraint missed by Keynes. (The same argument is reproduced in Jeffreys, 1939, page 167.) While I cannot judge of the level of novelty in Keynes' derivation with respect to earlier works, this derivation therefore produces a generic form of unidimensional exponential family, twenty-five years before their rederivation by Darmois (1935), Pitman

[8] As already discussed in footnote 5, the use of the MAP estimate reflects on the ambiguity at the time of the distinction between frequentist and Bayesian philosophies.

[9] Both terms "Fisher" and "estimation" are missing from the index.

[10] Keynes notes that *"differentiation assumes that the possible values of y* [meaning θ in our notations] *are so numerous and so uniformly distributed that we may regard them as continuous"* (page 196).

(1936) and Koopman (1936) as characterising distributions with sufficient statistics of constant dimensions. The characterisation of distributions for which the geometric and the harmonic means are MLEs then follows by a change of variables, $y = \log x$, $\lambda = \log\theta$ and $y = 1/x$, $\lambda = 1/\theta$, respectively. In those different derivations, the normalisation issue is treated quite off-handedly by Keynes, witness the function

$$f(x,\theta) = A\left(\frac{\theta}{x}\right)^{k\theta} e^{-k\theta}$$

at the bottom of page 198, which is not integrable *per se*. Similarly, the derivation of the log-normal density on page 199 is missing the Jacobian factor $1/x$ (or $1/y_q$ in Keynes' notations) and the same problem arises for the inverse-normal density, which should be

$$f(x,\theta) = Ae^{-k^2(x-\theta)^2/\theta^2x^2}\,\frac{1}{x^2}\,,$$

instead of $A\exp k^2(\theta-x)^2/x$ (page 200). At last, I find the derivation of the distributions linked with the median rather dubious because it does not seem to account for the non-differentiability of the absolute distance in every point of the sample. Furthermore, Keynes' general solution

$$f(x,\theta) = A\exp\left\{\int \frac{y-\lambda}{|y-\lambda|}\,\phi''(\lambda)\,\mathrm{d}\lambda + \psi(x)\right\}\,,$$

where the integral is interpreted as an anti-derivative, is such that the recovery of Laplace's distribution, $f(x,\theta) \propto \exp -k^2|x-\theta|$ involves setting (page 201)

$$\psi(x) = \frac{\theta - x}{|\theta - x|}\,k^2 x\,,$$

hence making ψ dependent on θ as well. In his summary, pages 204-205, Keynes (a) reintroduces a constant A for the normalisation of the density in the case of the arithmetic mean and (b) produces

$$f(x,\theta) = A\exp\left\{\phi'(\theta)\,\frac{\theta - x}{|\theta = x|} + \psi(x)\right\}$$

in the case of the median. This later form is equally puzzling because the ratio in the exponential is equal to the sign of $x-\theta$, leading to a possibly different ponderation of $\exp\psi(x)$ when $x<\theta$ and when $x>\theta$.

The method of least squares is also heavily attacked in Chapter XVII as *"surrounded by an unnecessary air of mystery"* (page 209), while conceding on the next page that it exactly corresponds to assuming the normal distribution on observations (a fact that is not correct either).

2.4 Keynes's [sketch of a] proposal for statistical inference

> *"I have experienced exceptional difficulty, as the reader may discover for himself in the following pages, both in clearing up my own mind about it*

> *and in expounding my conclusions precisely and intelligently."* ***A Treatise on Probability***, page 409.

Chapters XXXII and XXXIII advocates a very empirical approach to statistics that in a (very weak) way prefigures bootstrap, namely to derive the stability of a probability estimate by subdivising a *series* into a large enough number of *subseries* in order to assess the variability of the estimate or spotting heterogeneity. Keynes attributes the responsibility for this approach to Lexis and appears rather supportive of the latter, even though he comments that *"Lexis has not pushed his analysis far enough"* (page 401), before complaining about von Bortkiewicz, *"preferring algebra to earth"* (page 404). The description of the mechanism for dividing the series is not precisely clear (!), since it seems to depend on covariates: the sentence *"all conceivable resolutions into partial groups"* (page 395) is opposed to breaking *"statistical material into groups by date, place, and any other characteristic which our generalisation proposes to treat as irrelevant"* (page 397). The model thus constructed has a mixture flavour if the groups are made per chance, or a hierarchical one if not. Indeed, the description of *"the probability p for the group made up as follows"* (page 395)

$$p = \frac{z_1}{z} p_1 + \frac{z_2}{z} p_2 + \dots$$

clearly corresponds to a mixture, the z_i's being the component sizes. In any case, the description of Lexis' theory sums up as testing for variations between groups, i.e. by exposing a possible extra-binomial—called *supra-normal* by Keynes—variation. (He also mentions the possibility of an insufficient variation—the *subnormal* case—is attributed to dependence in the data, which *"cannot be handled by purely statistical methods"* (page 399)[11] Chapter XXXIII represents Keynes' attempts at summarising his own views about a constructive theory of statistical inference, but it sounds mostly like rephrasing Lexis' views, the main point being that one should work with *"*series of series *of instances"* (page 407) in order to check for the stability of the assumed model. The point that *"statistical induction is not really about the particular instance at all but a* series" (page 411) is valid but the attempt at checking that all subdivisions of the dataset show the same variability (*"until a* prima facie *case has been established for the existence of a stable probable frequency, we have but a flimsy basis for any statistical induction"*, page 415) is doomed when pushed to its extreme division into individual observations. *A Treatise on Probability* never explicitly derives a testing methodology in the sense of Gosset or of Fisher[12], despite mentions made of *"significant stability"* (pages 408 and 415). When discussing Lexis' dispersion in Chapter XXXII, Keynes refers to a case when *"the dispersion conformed approximately to the (...) normal law of error"* (page 358), but no entry is found on the contemporary Student's t tests or Pearson's

[11] The whole book considers handling dependent observations an impossible task, despite Markov's introduction of Markov chains a few years earlier. As pointed out by Aldrich (2008), the lack of feedback from Yule in *A Treatise on Probability* is apparent from the pessimistic views about dependent series, despite the proximity of the authors in Cambridge, since Yule had already engaged in building a statistical analysis of time series.

[12] Ronald Fisher also reviewed Keynes' book in 1923, concluding at the uselessness of Keynes' perspective on statistics.

χ^2 tests. The earlier criticisms of Keynes' about the extension of an observed model to future occurrences also apply in this setting, a fact acknowledged by the author: *"it is not conclusive and I must leave to others its more exact elucidation"* (page 419). Furthermore, the assessment of stability is not detailed and, while it seems to be based on normal approximations, the facts that the *same* data is used repeatedly and thus that the test statistics are dependent appear to be over-sighted by Keynes.

3 The Principle of Inverse Probability

> *"Bayes' enunciation is strictly correct and its method of arriving at it shows its true logical connection with more fundamental principles, whereas Laplace's enunciation gives it the appearance of* a new *principle specially introduced for the solution of causal problems."* **A Treatise on Probability***, page 175.*

When discussing the history of Bayes' theorem, Keynes considers in Chapter XVI that only Bayes got his proof right and that subsequent writers, first and foremost Laplace, muddled the issue (except for Markov). An interesting discussion revolves around the (obvious) fact that the prior probabilities of the different causes should be taken into account (*"the necessity in general of taking into account the* à priori *probability of the different causes"*, page 178). But one argument reflects the difficulty Keynes had with the updating of probabilities, as mentioned in the paragraph about the CLT: *"how do we know that the possibilities admissible* à posteriori *are still, as they were assumed to be* à priori*, equal possibilities* (page 176).[13]

3.1 Against the Principle of Indifference

> *"My criticism will be purely destructive and I will not attempt to indicate my own way out of the difficulties."* **A Treatise on Probability***, page 42.*

The *Principle of Indifference* is Keynes' renaming of the *Principle of Non-Sufficient Reason* advocated by Laplace and his followers for using uniform priors. Keynes (rightly) shows the inconsistency of this approach under (a) a refinement of the available alternative (pages 42-43) and (b) a non-linear reparameterisation of the model (page 45), the example being the change from ν into $1/\nu$. An extension of this argument on page 47 discusses the dependence of the uniform distribution on the dominating measure (although the book does not dally with measure theory), as illustrated by Bertrand's paradox. In the following and less convincing paragraphs, Keynes finds defaults with basic games examples where again the equidistribution depends on the reference measure.

> *"Who could suppose that the probability of a purely hypothetical event, of whatever complexity (...), and which has failed to occur on the one occasion*

[13] Note the accents used in à *priori* and à *posteriori*, although there are no accent in Latin. They may have stemmed from the way French writers first used those terms, even though *à posteriori* is also found in Jakob Bernoulli's *Ars Conjectandi*... The accent has vanished by the time of Jeffreys (1931).

> *on which the hypothetical conditions were fulfilled is no less than* $1/3$*?.”* **A Treatise on Probability**, *page 378.*

Similar arguments are advanced in Chapter XXX when debating about Laplace's law of succession.[14] Putting a uniform distribution on all possible alternatives is not coherent given that a subdivision of an alternative into further cases modifies the uniform prior. Further, a non-linear reparameterisation of a probability p into $q = p^n$ fails to carry uniformity from p to q. In concordance with the spirit of the time (Lhoste, 1923; Broemeling and Broemeling, 2003), the debate about whether or not the Principle of Indifference holds makes some sense, as shown by the subsequent defence by Jeffreys, but it does not hold much appeal nowadays because priors are recognised as reference tools for handling data rather than expressions of truth or of “objective probabilities”.

Chapter XXXI debates on the inversion of Bernoulli's Theorem, a notion that I interpret as Bayes formula applied to the Gaussian approximation to the distribution of an empirical frequency: on page 387

$$\frac{f(q')|h \cdot f(q)}{\sum f(q')|h \cdot f(q)}\,,$$

apparently meaning

$$\frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\,\mathrm{d}\theta}$$

in modern notations, is associated with the statement that *“all the terms can be determined numerically by Bernoulli's Theorem”*. Since this representation is somehow based on a flat reference prior (although I cannot understand the formula at the bottom of page 386 which seems to involve two distributions on the parameter θ), and thus on the Principle of Indifference, it is rejected by Keynes who cannot see a *“justification for the assumption that all possible values of q are* à priori *equally likely”* (page 387).

3.2 Against probabilising the unknown

> *“Laplace's theory requires the employment of both of two inconsistent methods.”* **A Treatise on Probability**, *page 372.*

The criticism of Bayesian (Laplacian's) techniques goes further than the rather standard debate about the choice of the prior. For Keynes, adopting a perspective that unknown probabilities can be modelled as random variables is beyond logical reasoning. Because an *unknown* probability is indeterminate, Keynes considers that *“there is* no *such value”* (page 373).[15] Therefore, the Bayesian notion of setting a probability distribution over the unit interval is both illogical and impractical, since *“if a probability is*

[14] Although no black swan glides in, the section contains the obligatory example of the probability of the sun rising tomorrow that found in almost every treatise on induction since Hume.

[15] A more positive perspective (see, e.g., Brady, 2004) is to consider that Keynes' stance prefigures the theory of imprecise probabilities à la Dempster–Schafer (see, e.g., Walley, 1991), as for instance when he states that *“many probabilities can be placed* between *numerical limits”* (page 160), but, to me, this

unknown, surely the probability, relative to the same evidence, of this probability has a given value, is also unknown" (page 373). Keynes then argues that, if the hyperprior probability is unknown, it should also be endowed with its own probability measure, inducing *"an infinite regress"* (page 373).

4 Conclusion

In conclusion, while Keynes' early interest in Probability and in Statistics is unarguable, *A Treatise on Probability* cannot be considered as a lasting contribution to Statistics, even from an historical perspective, given the immense developments taking place in Statistics at the time. It appears in the end as a scholarly exercise focussing on past books and lacking a vision of developments that would have made Keynes a statistician of his time, while the aggressive tone adopted towards most of the writers quoted in the book is undeserved when comparing the achievements of both sides. It is therefore no surprise the book has had no influence on the probability and statistics community: it would make no sense to advise students in the field to put aside major treatises to ponder through *A Treatise on Probability* as, to adopt Fisher's (1922) words, *"they would be turned away, some in disgust, and most in ignorance, from one of the most promising branches of mathematics."*

Acknowledgements

The author's research is partly supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy 75012 Paris) through the 2007–2010 grant ANR-07-BLAN-0237 "SPBayes". The first draft of this paper was written during the conference on Frontiers of Statistical Decision Making and Bayesian Analysis in San Antonio, Texas, March 17-20, held in honour of Jim Berger's 60th birthday, and the author would like to dedicate this review to him in conjunction with this event. He is also grateful to Eric Séré for his confirmation of the arithmetic mean distribution classification found in Chapter XVII.

References

Aldrich, J. 2008a. Keynes among the Statisticians. *History of Political Economy* 40(2): 265–316.

—. 2008b. R.A. Fisher on Bayes and Bayes' Theorem. *Bayesian Analysis* 3(1): 161–170.

Brady, M. 2004. *J.M. Keynes' Theory*. Xlibris Corporation.

---

mostly shows the same confusion between (interval) estimates and true probabilities found elsewhere in the book. The notion of replacing (pointwise) probabilities by intervals in the short Chapter XV is attributed to Boole—with another barb in the footnote of page 161—and it does not seem to be set to any implementable version in the statistical inference section (Part V).

Broemeling, L. and A. Broemeling. 2003. Studies in the history of probability and statistics XLVIII The Bayesian contributions of Ernest Lhoste. *Biometrika* 90(3): 728–731.

Burdzy, K. 2009. *The Search for Certainty*. Singapore: World Scientific.

Dale, A. 1999. *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*. 2nd ed. Sources and Studies in the History of Mathematics and Physical Sciences, Springer-Verlag, New York.

Darmois, G. 1935. Sur les lois de probabilité à estimation exhaustive. *Comptes Rendus Acad. Sciences Paris* 200: 1265–1266.

de Finetti, B. 1974. *Theory of Probability*, vol. 1. New York: John Wiley.

Fienberg, S. 2006. When did Bayesian inference become "Bayesian"? *Bayesian Analysis* 1(1): 1–40.

Fisher, R. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A* 222: 309–368.

—. 1923. Review of J.M. Keynes's Treatise on Probability. *Eugenic Review* 14: 46–50.

Hald, A. 1999. On the history of maximum likelihood in relation to inverse probability and least squares. *Statist. Sci.* 14(2): 214–222.

Jeffreys, H. 1922. The Theory of Probability. *Nature* 109: 132–133.

—. 1931. *Scientific Inference*. 1st ed. Cambridge: The University Press.

—. 1939. *Theory of Probability*. 1st ed. Oxford: The Clarendon Press.

Keynes, J. 1911. The Principal Averages and the Laws of Error which Lead to Them. *J. Royal Statistical Society* 74: 322–331.

—. 1921. *A Treatise on Probability*. London: Macmillan and Co.

Koopman, B. 1936. On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* 39: 399–409.

Lhoste, E. 1923. Le Calcul des Probabilités Appliqué à L'artillerie 91: 405–423, 516–532, 58–82 and 152–179.

Lindley, D. 1968. John Maynard Keynes: Contributions to Statistics. In *International Encyclopedia of the Social Sciences*, vol. 8, 375–376. New York: Macmillan Company & The Free Press.

Pitman, E. 1936. Sufficient statistics and intrinsic accuracy. *Proc. Cambridge Philos. Soc.* 32: 567–579.

Robert, C., N. Chopin, and J. Rousseau. 2009. Theory of Probability revisited (with discussion). *Statist. Science* 24(2): 141–172 and 191–194.

Stigler, S. 2002. Statisticians and the history of Economics. *Journal of the History of Economic Thought* 24(2): 155–164.

Taleb, N. 2008. *Fooled by Randomness*. 2nd ed. Random House.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probability*. New York: Chapman and Hall.